Human Capital Growth in a Cross Section of US Metropolitan Areas

Authors	Christopher H. Wheeler
Working Paper Number	2005-065A
Creation Date	
Citable Link	https://doi.org/10.20955/wp.2005.065
Suggested Citation	Wheeler, C.H., 2005; Human Capital Growth in a Cross Section of US Metropolitan Areas, Federal Reserve Bank of St. Louis Working Paper 2005-065. URL https://doi.org/10.20955/wp.2005.065

Published In	Federal Reserve Bank of St. Louis Review

Human Capital Growth in a Cross Section of US Metropolitan Areas

Christopher H. Wheeler*
Research Division
Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166

September 2005

christopher.h.wheeler@stls.frb.org

Abstract

Human capital is typically viewed as generating a number of desirable outcomes, including economic growth. Yet, in spite of its importance, few empirical studies have explored why some economies accumulate more human capital than others. This paper attempts to do so using a sample of more than 200 metropolitan areas in the United States over the years 1980, 1990, and 2000. The results reveal two consistently significant correlates of human capital growth, defined as the change in a city's rate of college completion: population and the existing stock of college-educated labor. Given that population growth and human capital accumulation are both positively associated with education, these results suggest that the geographic distributions of population and human capital should have become more concentrated in recent decades. That is, larger, more-educated metropolitan areas should have exhibited the fastest rates of increase in both population and education and thus 'pulled away' from smaller, less-educated metropolitan areas. The evidence largely supports this conclusion.

JEL: J24, R11, R12, R23

Keywords: Human Capital, Urban Growth, Local Labor Markets

---

* The views expressed herein are those of the author and do not necessarily represent the official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

1 Introduction

Human capital is now commonly held to be one the fundamental drivers of economic growth. To be sure, the notion that the skills possessed by an economy's workforce contribute positively to technological change and the growth of productivity is an intuitively appealing one. Yet, there is also a fair amount of empirical evidence that lends it support. In particular, a sizable literature in the last two decades has established a strong statistical association between human capital (usually captured by educational attainment) and the growth of employment, productivity, and income. What is more, this relationship holds with striking regularity at different levels of geographic aggregation, including countries (Barro (1991)), US states (Barro and Sala-i-Martin (1992)), cities and metropolitan areas (Glaeser et al. (1995), Glaeser and Saiz (2003), Simon and Nardinelli (2002)).

Economic growth, however, is only one benefit that has been associated with human capital. A variety of studies also suggest that greater educational attainment within local economies (e.g. states or cities) may tend to be accompanied by lower rates of crime (Lochner and Moretti (2004)), greater civic involvement (Dee (2004), Milligan et al. (2004)), and less political corruption (Glaeser and Saks (2004)). Clearly, because these are desirable outcomes, identifying the determinants of human capital growth is a worthwhile undertaking. Unfortunately, while a host of theoretical models have done so[1], surprisingly little empirical research has followed suit. Most existing studies have focused on what human capital produces rather than why some economies accumulate more of it than others.[2] As such, our understanding of human capital accumulation remains limited.

This paper seeks to address this matter by looking at the growth of human capital in a sample of more than 200 US metropolitan areas identified in the decennial US Census over the years 1980, 1990, and 2000. Defining human capital accumulation as

[1] See Barro and Sala-i-Martin (1995) for a survey of human capital-based models of growth.

[2] There are two notable exceptions. Moretti (2004) offers a short analysis of the determinants of changing college attainment rates among US metropolitan areas, similar to what I do here. Glaeser and Saiz (2003) examine whether educational attainment responds to economic growth. With both of these papers, however, the primary issue under consideration is *not* the determinants of human capital growth. Consequently, their analyses are much more cursory with respect to this issue than what is presented here.

the change in the fraction of a metropolitan area's employed labor force with a bachelor's degree or more, I find that metropolitan areas with larger populations and higher fractions of their workers with a bachelor's degree tend to accumulate human capital at faster rates than less-populous, less-educated metropolitan areas. The results suggest that a 1-standard deviation increase in either total resident population or the fraction of workers with a 4-year college degree (in the cross section of metropolitan areas) tends to be associated with a 0.4 to 0.7 percentage point rise in the share of college graduates in the workforce over the next decade. These estimated magnitudes, it should be noted, are *not* meant to be interpreted as causal, but simply to quantify the strength of the observed associations between these two variables and the accumulation of highly educated workers. Although some evidence suggests that certain measures of industrial composition and observable city-level amenities (e.g. restaurants and schools) are also associated with changes in the college fraction, none are as robustly correlated as population and the existing level of human capital.

Intriguingly, these findings seem to suggest that the geographic distribution of human capital across the cities of the US should have grown more concentrated (or unequal) between 1980 and 2000. After all, because human capital accumulation tends to be positively associated with the current level of human capital, the gap between initially high-education cities and low-education cities ought to have widened in recent decades. The evidence strongly supports this conclusion. Various measures that characterize the degree of spread in the distribution of metropolitan area-level college attainment show rising dispersion between 1980 and 2000.

In addition, because previous research has established a positive link between population growth and education (e.g. Glaeser et al. (1995)), one would expect to find a similar pattern of 'divergence' in population levels across US metropolitan areas in recent decades. That is, if more populous cities accumulate highly educated workers more quickly than less populous ones, then they should also gain population faster too. Rising educational attainment fuels population growth which, in turn, spurs human capital accumulation, and so on. This conclusion is also largely borne out in the data. The distribution of the logarithm of population became more concentrated between 1980 and 2000.

Although one might surmise that rising concentrations of population and education in the largest and most-educated cities have also led to a greater concentration of income, the evidence on this issue is somewhat mixed. In particular, while the data show that the distribution of metropolitan area-level average log hourly wages grew wider between 1980 and 1990, they also show that it narrowed slightly between 1990 and 2000. Growing concentrations of population and college-educated workers in the largest and most human capital-abundant metropolitan areas, then, have not been accompanied by substantial increases in the degree of inter-city (average) earnings inequality.

2 Data

The data used in the analysis are taken primarily from the 5 percent public use samples of the 1980, 1990, and 2000 US Census as reported by the Integrated Public Use Microdata Series (Ruggles et al. (2004)). These data files include a variety of personal characteristics, including age, education, and earnings for samples of more than 11 million individuals in each year, as well as information about each individual's place of residence. These data are used to construct a time series of metropolitan area-level characteristics, including human capital.

In principle, 'human capital' could be defined in many different ways: e.g., time spent on a particular job, time spent working on all jobs, numbers of different jobs held, educational attainment, some measure of 'innate' ability or productivity. This paper takes a standard approach by using educational attainment, which can be justified by noting that (*i*) schooling has been shown to have a significant causal influence on individual productivity, at least as quantified by earnings (Card (1999)), and (*ii*) it tends to be strongly correlated with a variety of outcomes commonly theorized to be tied to 'human capital,' including growth. For these reasons, education is treated as a suitable metric for human capital. More specifically, I use the fraction of a metropolitan area's employed labor force with a bachelor's degree or more since previous work on economic growth and education externalities in cities have found this particular quantity to capture variation in educational attainment reasonably well.[3]

[3] See, for example, Black and Henderson (1999) and Moretti (2004).

Formally, metropolitan areas in the analysis represent either metropolitan statistical areas (MSAs), New England County Metropolitan Areas (NECMAs), or Consolidated Metropolitan Statistical Areas (CMSAs) in the event that an MSA or NECMA belongs to a CMSA.[4] A total of 210 of these local markets are identified in the 1980 data, 206 in 1990, and 245 in 2000. Only 188 appear in all three Census years.

Additional characteristics describing metropolitan areas are derived from the USA Counties on CD-ROM (US Bureau of the Census (1999)) and from County Business Patterns (CBP) files for the years 1980, 1990, and 2000. The former data set provides information about county-level population and land area which is used to generate population and population density figures at the metropolitan area level.[5] The latter reports the numbers of various types of private sector establishments (e.g. restaurants and bars) which are used to characterize the amenity-value of a metropolitan area. Further details about the data appear in the Appendix.

3 Empirical Findings

3.1 Human Capital and Urban Agglomeration

Within the U.S., human capital has typically been concentrated in metropolitan areas. Among workers in the Census samples used here, 86.1 percent of all college graduates resided in a metropolitan area in 1980. By 2000, this figure had risen to 89.9 percent. In contrast, approximately 78 percent of workers with only a high school diploma were metropolitan dwellers in either year.

Why are highly educated workers drawn to cities? Numerous characteristics, of course, distinguish metropolitan areas from non-metropolitan areas and, thus, could offer some semblance of an answer. Besides larger and better-educated populations, urban agglomerations also tend to possess greater numbers of industries that highly educated workers may find particularly appropriate or appealing given their skills (e.g.

---

[4] Throughout the paper, I use the terms 'metropolitan area' and 'city' interchangeably for expositional purposes. In all cases, local markets refer to MSAs, NECMAs, or CMSAs.

[5] County-level population data for the year 2000 is derived from the population estimates program of the US Census Bureau at www.census.gov/popest/estimates.php. In all years, land area from 1990 is used to compute density.

professional and technical services). Metropolitan areas also tend to offer a greater array of amenities (e.g. restaurants and museums) which may serve to attract and maintain a pool of highly educated labor (see Glaeser et al. (2001)).

Economically, the estimated returns to education do tend to be particularly high in metro areas. Consider, for instance, the results from a regression of log hourly earnings on 5 educational attainment indicators (no high school, some high school, a high school diploma only, some college or an associate's degree, a bachelor's degree or more), 8 indicators representing years of potential work experience[6], a metropolitan residence dummy, and interactions between metro residence and each of the education and experience variables.[7] To keep the analysis simple, I have limited the sample of workers used for this regression to white males between the ages of 18 and 65. I have also performed the estimation separately for the 1980 and 2000 samples to account for any changes in the coefficient values over time.[8]

The resulting coefficient estimates, which for the sake of conciseness have been limited to the education variables, appear in Table 1. The raw coefficients on the five educational attainment dummies in the first 5 rows of results can be interpreted as the average log wages (conditional on all of the other covariates in the model) for workers in these education groups who reside *outside* of a metro area. The average log wages for workers inside metro areas is then given by the sum of these raw coefficients and the corresponding interaction listed in the remaining rows of the table.

With this interpretation in mind, it is evident that, although college graduates earn more than workers with less schooling, the premium associated with a college degree is particularly high within metropolitan areas. In non-urban areas, for example, college-educated workers earned approximately 30 percent more than workers with only a high school diploma in 1980.[9] Within metropolitan areas, that differential was 45 percent. By

---

[6] These indicators represent 6-10 years of experience, 11-15 years, 16-20 years, 21-25 years, 26-30 years, 31-35 years, 36-40 years, and 41 or more years.
[7] The regressions also include dummies for marital status, disability, veteran's status, and foreign-born status.
[8] The 1990 5 percent sample does not report metropolitan status for all individuals in the sample. Hence, estimating the regression for this year is not possible.
[9] Percentages are derived from the estimates in Table 1 by exponentiating the log wage differential and subtracting 1. A 26 log point differential between college and high school graduates in non-metro areas in 1980, for example, corresponds to roughly 30 percent.

the year 2000, the college premium had risen to 49 percent outside of metro areas, 75 percent within them. In terms of raw (conditional) wage levels, college graduates earned an average of $10.48 per hour outside of metro areas in 1980, $12.26 within them.[10] By 2000, these figures stood at respectively $10.80 and $13.40, implying a 20-year growth rate of roughly 17 percent in rural areas, but 24 percent in urban areas.

These figures, of course, should not be interpreted causally. That is, a highly educated worker's metropolitan status does not necessarily *cause* her to earn more than if she were situated in a smaller labor market. On the contrary, the results may reflect, at least in part, a selection mechanism by which the most productive, highly educated workers have chosen to live in cities. Still, these results seem to suggest that there are strong economic incentives for highly educated workers to reside in urban areas.

To gain a better sense of which factors (e.g. metropolitan area size, existing human capital, education premia, industrial composition) may underlie human capital accumulation, I now turn to the analysis of a cross section of metropolitan areas. The underlying goal is to exploit the variation exhibited across cities with respect to their education, size, and other characteristics to draw inferences about which features are most strongly associated with the growth of human capital.

3.2 Correlates of Human Capital Accumulation: Baseline Results

As noted previously, the Census data used in this paper identify more than 200 metropolitan areas in each of the three years (1980, 1990, 2000) considered. Using this sample, I estimate the following simple regression in which the change in metro area *i*'s college fraction during decade *t*, $\Delta Coll_{i,t}$, is specified as

$$(1) \quad \Delta Coll_{i,t} = \mu + \delta_t + \beta X_{i,t} + \varepsilon_{i,t}$$

where μ is a constant; δ_t is a decade-specific fixed effect; $X_{i,t}$ is a set of characteristics describing the metropolitan area at the *beginning* of the decade; and $\varepsilon_{i,t}$ is a stochastic

[10] These estimates are based upon exponentiating the coefficients in Table 1.

element, assumed to be uncorrelated across metropolitan areas but potentially correlated within them (i.e. $\varepsilon_{i,t}$ and $\varepsilon_{i,s}$ may show some non-zero association). This equation is meant to be analogous to those used in empirical studies of economic growth in which a measure of growth is regressed on a set of initial characteristics (e.g. Barro (1991), Glaeser et al. (1995)).

Among the characteristics considered in the vector $X_{i,t}$ are the following: (*i*) an estimate of a metropolitan area's return to a college degree[11], (*ii*) its level of human capital (given by the fraction of college-educated workers in the labor force), (*iii*) its raw size (given by the logarithms of population and population density), and (*iv*) its broad industrial composition (measured by shares of total employment accounted for by each of 20 industries). Summary statistics for each of these regressors appear in Table 2.[12]

Results are given in Table 3. The first column, labeled *I*, reports the resulting coefficients when each covariate is entered into the regression separately. In all instances, estimation of equation (1) also includes a set of three region dummies to account for any exogenous differences in the rate of human capital accumulation in different parts of the country.[13]

Based on the estimates, many of these regressors do turn out to be significantly associated with the growth of the college fraction, at least in a simple, univariate sense. Metropolitan areas with initially larger populations, higher levels of population density, and larger fractions of workers with a bachelor's degree or more all see their college attainment rates rise by more over the following decade than smaller, less-dense, less-educated metropolitan areas. In addition, greater fractions of employment accounted for by industries such as agriculture, mining, and manufacturing (either durable or non-durable) tend to correlate negatively with human capital accumulation while a strong presence of industries like finance, insurance, real estate and business and repair services, are positively associated with the change in the college attainment rate. Given that the

[11] Metropolitan area college degree returns are derived from city-year specific regressions of log hourly wages on 5 education indicators, 8 experience indicators, and dummies for marital status, disability, veteran's status, and foreign-born status. The coefficient on the college completion dummy is used to estimate the return to a college degree.

[12] Because they are easier to interpret, Table 2 lists summary statistics for population and population density levels rather than logarithms. In the regression analysis, I utilize these variables in log form which is reasonably standard in the empirical literature on cities.

[13] A list of the state-level composition of the four US Census regions appears in the Appendix.

former set of industries tends to employ fewer highly educated workers than the latter set of industries (see Table 4), these associations are rather intuitive. The estimated city-specific return to a college degree, while positive, is not statistically important. Greater discussion of this last regressor is provided below.

The next two columns of results, *II* and *III*, report the coefficients from two different specifications of (1) in which various combinations of these covariates appear. The longer of these (*III*) suggests that, unlike what is reported above, very few of the initial industry shares are significantly associated with human capital accumulation. Indeed, comparing the results from columns *I* and *III*, only one industry share enters significantly in both cases: finance, insurance, real estate. Industrial composition, therefore, seems largely unimportant for explaining the growth of human capital, at least once we have conditioned on initial education, size, and returns.

Among the remaining covariates, only two show consistently positive and significant associations with human capital accumulation: log population and the initial college fraction. Both of these regressors produce significant coefficients in all three reported specifications. Log density, by contrast, becomes insignificant when industry shares are included, and the initial return to a college degree enters negatively (and significantly) in specifications *II* and *III*. This latter result may simply reflect the inverse association between various measures of urban growth (e.g. population and average earnings) and initial wages, which is a common finding in the urban economics literature (e.g. Glaeser et al. (1995)). Higher returns to a college degree, not surprisingly, tend to be associated with higher average wages overall in these data. As growth slows, human capital accumulation tends to slow as well.[14]

How significant are the estimated associations between, on the one hand, initial log population and the college completion rate and, on the other, the subsequent change in the college completion fraction? Based on the point estimates from the longest specification in Table 3, a 1-standard deviation increase in log population (in the cross section) corresponds to a 0.43 percentage point rise in the college attainment rate over the

[14] The positive coefficient on the initial estimated college return in specification *I* may therefore emanate from omitted variable bias. As shown in Section 3.1, returns to a college degree tend to be higher in metro areas, suggesting a positive association with population and the college attainment rate. Not including these two variables in specification *I* may therefore bias a truly negative coefficient on initial returns upward.

next decade. A 1-standard deviation increase in the initial fraction of workers with a bachelor's degree or more has a somewhat larger implied association: a 0.72 percentage point rise in the college attainment rate over the next 10 years.[15] Although they may seem small compared to average college completion rates near 22 percent for the metropolitan areas in the sample, these magnitudes are far from negligible. In particular, they represent between 20 and 34 percent of the cross sectional standard deviation of the 10-year change in the college fraction in these data, which is approximately 2.1 percentage points.

3.3 Robustness

In this section, I consider a few simple alterations to the statistical analysis to assess the robustness of the results. The first seeks to account for the influence of certain amenities (e.g. restaurants, theaters, museums) on human capital accumulation. As noted previously, Glaeser et al. (2001) have demonstrated that cities have significant consumption aspects which seem to influence the willingness of individuals to live in dense urban environments. If the highly educated have an especially strong preference for these characteristics, amenities may play an important part in human capital accumulation that the analysis above misses. Indeed, it may not be a city's population or initial level of educational attainment that are important for explaining the growth of a city's college share, but its array of urban amenities. Population or education may simply be proxies for these types of characteristics. To explore this possibility, I consider the influence of the following 7 amenities: eating and drinking establishments; movie theaters; live entertainment venues; museums, botanical gardens, zoos; elementary schools; colleges and universities; hospitals.[16] Initial values of these quantities, expressed in per capita terms, are added to equation (1).

The second alteration takes a different approach to controlling for the influence of industrial composition. While initial shares of a metropolitan area's employment across a

[15] The cross-sectional standard deviations for log population and the college completion rate are roughly 1.08 and 0.065.

[16] Many of these variables were identified by Glaeser et al. (2001) as being significantly related to population growth.

broad array of sectors may offer some explanatory power with respect to human capital accumulation, how they change over time may be more relevant. That is, it may not be the initial share of employment in a city's durable manufacturing sector that affects its college fraction, but the *change* in the fraction accounted for by that sector. Again, as demonstrated in Table 4, there are substantial differences in college attainment across the 20 industries considered. One might, therefore, expect that rising shares of employment in, say, retail trade, which employs relatively few college-educated workers, to have a negative influence on a city's overall level of education, whereas a rise in the fraction of workers employed in educational services, which employs primarily college-educated labor, would accomplish just the opposite. To address this potential misspecification of the regression, I include contemporaneous changes in each sector's employment share in (1) and drop the initial levels.

Results appear in Table 5. As before, coefficient estimates from three different specifications are reported to gauge the sensitivity of the findings to variations in the model. The first column, labeled *I*, reports coefficients from the regression of the change in the college attainment rate on the initial estimated return earned by college graduates, log population, log density, the initial college fraction, and initial per capita quantities of the 7 amenities listed above.[17] Interestingly, three of these amenities enter significantly – eating and drinking places, elementary schools, and hospitals – although the first two do so positively whereas the third does so negatively.[18] In spite of this result, however, none of the remaining coefficients, including those on log population and the college fraction, change appreciably from what was reported above.

The second column of results drops these 7 amenities and adds changes in 19 of the 20 industry employment shares to determine whether specifying industrial composition in 10-year differences rather than initial levels makes any difference in the remaining coefficient estimates.[19] Compared to the specification of industry mix in initial levels, a greater number of industries now produce significant associations, and

[17] Results were similar when the 20 initial industry shares were included. Since reporting all of these additional coefficients would have been excessive, I have omitted them from the regression.
[18] The number of hospitals per capita may be associated with the growth in the numbers of relatively old workers who tend to possess less education than younger workers
[19] Because changes in all 20 industry shares (by definition) sum to 0, I drop the change in the employment share of agriculture, forestry, and fisheries.

many of these are quite reasonable. An increase in the importance of finance, insurance, real estate as well as social and business and repair services, for example, should be associated with increases in the fraction of workers with a bachelor's degree or more. These sectors, after all, tend to employ relatively large proportions of college-educated labor. This conclusion is indeed borne out regardless of whether the 7 amenities listed above are included in the regression (column *III*) or not (column *II*).

At the same time, inclusion of changes in industrial composition has very little impact on the estimated initial population and college fraction coefficients. Both remain statistically significant, and the magnitudes are very similar to those reported in all previous specifications. Such a finding only reinforces the conclusion that, even after accounting for a city's industrial composition, a city's initial scale and education are strongly associated with the rate at which it accumulates highly educated workers.

Of course, characterizing the industrial composition of a metropolitan area using a set of 20 broad sectors is less than ideal. There is a fair amount of heterogeneity inherent in each industry; hence, this classification scheme may miss important differences in the types of employers present in each metropolitan area. For example, the types of employers belonging to the non-durable manufacturing sector in one city (e.g. drugs or chemicals) may be quite different from those in another (e.g. textiles or food processing). These differences may be important in explaining the growth of human capital, but would be missed by the present analysis. More seriously, these unmeasured differences may very well be directly correlated with either population or the college fraction. In such an instance, the coefficients reported thus far for these two regressors would be upwardly biased. [20]

I attempt to address this matter by looking, instead, at a collection of more than 200 industries, representing sectors at a mostly three-digit (standard industrial classification) level, although some two- and four-digit industries, as well as

[20] For example, one city may attract human capital because it has a strong presence of non-durable manufacturing which hires mostly highly educated workers (e.g. drugs and chemicals) whereas another may attract less human capital because it has a strong presence of non-durable manufacturing which hires primarily less-educated workers (e.g. textiles and food processing). The presence of high- and low-human capital non-durable manufacturers will therefore be directly related to each city's initial stock of human capital, but the association between industrial composition and human capital accumulation (which is real in this example) will be picked up by the initial stock of human capital.

combinations of two-, three- and four-digit industries, also appear.[21] These are the most detailed industrial categories available in the decennial Census files.

Unfortunately, because adding more than 200 industry shares to the estimation of (1) is infeasible from a practical perspective, I use the following approach. First, I create a 'predicted' college attainment fraction, $PColl_{i,t}$, for each metropolitan area, *i*, in each year *t*, as follows:

$$(2) \quad PColl_{i,t} = \sum_{s=1}^{N_{i,t}} Share_{s,i,t} Coll_{s,t}$$

where $Share_{s,i,t}$ is the share of sector *s* in metropolitan area *i*'s total employment in year *t*, $Coll_{s,t}$ is the college completion fraction for sector *s* in year *t* (calculated using aggregate data for the US), and $N_{i,t}$ is the number of sectors in metropolitan area *i* in year *t*. Second, I compute a 'residual' college fraction given by $(Coll_{i,t} - PColl_{i,t})$ which measures the difference between a city's actual college-completion fraction and the fraction that would result if its industries resembled the national average. I interpret this difference as the part of a city's college-attainment fraction which is not explained by its detailed industry composition. I then consider regressions of the form

$$(3) \quad \Delta(Coll_{i,t} - PColl_{i,t}) = \mu + \delta_t + \beta X_{i,t} + \varepsilon_{i,t}$$

where two specifications of the regressors $X_{i,t}$ are considered: one which controls for the estimated college return, log population, log density, and the college fraction, all in initial levels; the other which further adds initial values of the 7 amenities discussed above. The resulting estimates appear in Table 6.

---

[21] Specifically, there are 223 industries in the 1980 data, 221 in the 1990 data, and 214 in the 2000 data. These are identified using consistent codes established using the correspondence provided by the U.S. Bureau of the Census. Tobacco and crude petroleum and natural gas are examples of two-digit industries; drugs, electric light and power, and grocery stores are examples of three-digit industries; jewelry stores and retail florists are examples of four-digit industries.

In general, they demonstrate very little change from what has already been reported. Among the amenities, the same three variables – eating and drinking places, elementary schools, and hospitals per capita – all enter significantly and with the same signs as before. Additionally, the initial college-return produces a significantly negative coefficient, while the logarithm of population and the initial fraction of college-educated workers in total employment generate significantly positive coefficients.

With these latter two regressors, it is worthwhile noting that the coefficients are now somewhat smaller than what is reported in Tables 3 and 5. For example, in the longer of the two specifications in Table 6, log population produces a coefficient of 0.003 rather than 0.004 previously, whereas the initial college completion rate generates a coefficient of 0.07 rather than 0.11. These decreases are consistent with the idea mentioned previously that using 20 broad industry shares leads to upwardly biased coefficients on the initial college fraction and log population. Still, the evidence is remarkably consistent with respect to the influence of population and education. Regardless of how the statistical model is specified, these two covariates are significant predictors of human capital accumulation.

3.4 Human Capital, Growth, and Divergence

The finding that more populous and educated cities tend to experience the largest increases in human capital has an intriguing implication with respect to the geographic distributions of population and college-educated labor. Specifically, it suggests that the distributions of these two quantities should have been characterized by increasing concentration over the 1980-2000 period. Human capital accumulation, after all, tends to be faster in cities with larger fractions of highly educated workers initially. This mechanism should then lead to a growing gap between the education levels across cities over time as the top end of the distribution pulls away (or 'diverges') from the bottom. Because previous work has shown that more educated cities also tend to see faster population growth (e.g. Glaeser et al. (1995)) we arrive at a similar implication with respect to the distribution of population. This section examines whether there has been this type of 'divergence' in the distribution of these two quantities.

Before doing so, I begin by establishing some basic results relating the growth of two quantities – population and average hourly wages – to education. While the former is of greater interest in this particular exercise, the latter more closely resembles the object of interest in most studies of economic growth (i.e. per capita income). Results from the regression of each quantity's 10-year growth rate on the initial level of human capital appear in the specifications labeled *I* in Table 7.[22] Not surprisingly, each shows a significantly positive association with initial education. Here, the magnitudes indicate that a 1-standard deviation (i.e. a 6.5 percentage point) increase in a city's college attainment rate tends to be accompanied by a 1.8 percentage point rise in its rate of population growth and a 1.7 percentage point rise in its rate of average wage growth over the next 10 years. These figures represent, respectively, 16 and 20 percent of the cross-sectional standard deviations in these two growth series. These associations, therefore, seem to be both statistically and economically important.

To explore whether there has been divergence across city-level human capital, population, and average wages, I consider two approaches. The first looks for so-called β-convergence, the test for which involves a simple regression of the growth of a quantity on its initial level.[23] A negative coefficient on the initial level of a variable would indicate a tendency for that quantity to converge to a common level across metropolitan areas. After all, a negative coefficient would indicate that cities with low levels of human capital, for example, would experience faster human capital growth than cities with high levels. This process should generate a less-concentrated distribution of human capital over time as the bottom of the distribution catches up with the top. The second approach looks for σ-convergence which is based upon how the cross sectional dispersion of a particular quantity changes over time. Decreasing dispersion (i.e. falling concentration) would be indicative of σ-convergence.[24]

---

[22] As with all of the other regressions, these include three region dummies and an indicator for the 1980-90 decade.

[23] Again, all regressions also include three region dummies and a time effect to pick up differences in growth across decades. The β in β-convergence refers to the coefficient on the initial level of a variable in a growth regression.

[24] The σ in σ-convergence refers to the standard deviation which is commonly represented in statistics by the Greek letter σ. Barro and Sala-i-Martin (1995) provide an overview of the statistical techniques commonly used in studies of convergence/divergence.

The β-convergence results for metropolitan area college attainment are already well-established in the findings shown thus far. The strong positive association between the initial level of a city's college fraction and its subsequent change over the next decade indicates divergence in this variable. Results for the logarithm of population and the average log hourly wage appear in the specifications labeled *II* in Table 7.

The population series also shows divergence which, intuitively, is precisely what one would expect in light of the results shown to this point. Larger populations tend to be associated with more rapid human capital accumulation which raises education levels. This, in turn, leads to faster population growth. Hence, one would expect to see a positive association between initial population and its subsequent rate of growth. Interestingly, however, the positive association between initial population and its subsequent growth also holds after conditioning on the initial college fraction and the initial average log hourly wage. This result is reported in specification *III*. The direct association between population and population growth, therefore, does not seem to be driven entirely by education. There is some aspect of metropolitan area size that, independent of education, draws additional population.

Average hourly wages, by contrast, show evidence of convergence rather than divergence. That is, higher average wages tend to be followed by slower rates of wage growth over the next decade. This finding too is sensible given the evidence already presented. Recall, higher wages tend to be accompanied by slower human capital accumulation subsequently. The significantly negative coefficients on the initial college return in the regression results presented above demonstrate this point clearly. Slower human capital accumulation, then, implies slower growth of average hourly wages. Thus, one would expect to see a negative association between initial average wages and future wage growth. This relationship turns out to hold whether initial education and log population are accounted for or not (compare specifications *II* and *III*).

To look at σ-convergence, I need a measure that characterizes the degree of spread in the distributions of human capital, log population, and the average log hourly wage.[25] In an effort to keep the analysis broad, I consider several possibilities: the

[25] For this exercise, I use population and average wages in logarithmic form because the distributions of their levels will tend to show increasing dispersion even if growth is unrelated to the initial level. For

standard deviation and a host of inter-quantile differences (e.g. the difference between the 90th percentile and the 10th). One important consideration in looking at these distributional features is maintaining a consistent sample of metropolitan areas. The sample of metropolitan areas identified by the Census does change from one year to the next. As a consequence, there may be changes in the degree of spread in the distribution of these variables that stem from changes in the composition of the sample rather than an actual convergence or divergence mechanism. In computing these distributional features, then, I confine the sample to those 188 metropolitan areas that appear in all three years.

The resulting estimates appear in Table 8. Looking at the distribution of college attainment rates, it is evident that, although there has been an increase in the fraction of workers with a bachelor's degree or more at all points of the distribution, that increase has been larger at the top than at the bottom. The 90th percentile, for example, rose by more than 10 percentage points between 1980 and 2000, increasing from 0.238 to 0.339. The corresponding increases for the median and 10th percentiles over this period were 7.3 and 4.8 percentage points. Accordingly, each of the four listed percentile gaps (90-10, 90-50, 50-10, 75-25) grew wider over time. Rising dispersion can also be inferred from the evolution of the standard deviation which started at 0.043 in 1980, rose to 0.054 in 1990, and stood at 0.064 by 2000. Evidently, human capital became more unevenly distributed during this time frame.[26]

The logarithm of population, the distributional features of which appear just below the human capital results in Table 8, reveals a similar trend. On average, metropolitan areas in the US experienced population gains between 1980 and 2000, and these gains were registered at all five quantiles of the distribution. Again, however, the gains tended to be somewhat larger at the top of the distribution than at the bottom. With the exception of the inter-quartile difference (75-25) which did not change between 1980 and 2000, all other quantile differentials increased in both decades. Increasing dispersion in the logarithm of population also shows up in the standard deviations, which increased

example, the gap between the populations of two cities, one with population of 100, the other with a population of 1000, will grow wider if both cities grow by the same percentage (and possibly if the smaller city grows by a larger percentage).

[26] Moretti (2004) documents a similar rise in the degree of human capital 'inequality' across US metropolitan areas.

from 1.057 in 1980 to 1.075 in 1990 to 1.092 in 2000. Log population, therefore, also shows evidence of both types of divergence.

These results are particularly striking because they stand somewhat at odds with what conventional economic analysis might suggest. Indeed, as cities grow in population, they tend to become more congested which, in turn, raises costs (financial and otherwise) to both workers and employers. These 'agglomeration diseconomies' should, therefore, work to slow subsequent rates of population growth as firms and workers seek less congested labor markets. Similarly, as the fraction of workers with a bachelor's degree rises, the relative return received by college-educated workers (all else equal) should decline since the supply of such workers has risen relative to demand. This is a standard diminishing marginal productivity argument whereby the return received by a factor of production (e.g. college-educated labor) declines as it is used more intensively relative to all other inputs. A lower return paid to college-educated labor, of course, should reduce the rate at which workers with a bachelor's degree move into an area. Empirically, however, there seems to be little support for these theoretical ideas in the data.

Average hourly wages, recall, show a very different pattern. Regressions of wage growth on initial wage levels reveal a significantly negative relationship between the two. One might expect, therefore, to see a decrease in the degree of dispersion in the distribution of city-level average log wages. The estimated dispersion measures in Table 8, however, only show a decrease between 1990 and 2000 when the standard deviation and all four quantile differences narrowed. During the 1980s, all but the 50-10 difference increased. These results demonstrate an important difference between β- and σ-convergence. Although a negative association between the initial level of a variable and its subsequent growth rate may certainly reduce the degree of variance in a distribution, it may also increase it. Durlauf and Quah (1999), for example, show how β-convergence may generate a wider distribution if economies with low levels of a variable 'overshoot' economies with high levels. Distributional dynamics of this sort may help to explain these results.

Another possible explanation may relate to the influence of population and the college fraction, which, as shown in Table 7, tend to be positively related to wage growth over the next decade. As these two variables have diverged, they may have led to a divergence in wage levels during the 1980s if their influence outweighed the natural tendency for wage levels to converge. Assuming that this natural tendency was stronger during the 1990s, of course, the cross-sectional dispersion in average log wages would have declined.[27]

4 Concluding Discussion

This paper has explored the issue of human capital accumulation across a set of US metropolitan areas. Among the more prominent findings is that cities with larger populations and larger fractions of workers with college degrees tend to see faster growth in their stocks of human capital. Because human capital also tends to be positively associated with population growth, this process has led to a divergence of both human capital and population in the US between 1980 and 2000. Hence, the largest and most educated cities in the country have tended to accumulate population and human capital faster than smaller, less-educated cities.

The divergence of human capital and population has not, interestingly, generated much divergence with respect to average wage levels across cities. Although the amount of dispersion in the distribution of metropolitan area-level average wages did grow larger between 1980 and 2000, this growth was experienced during the decade of the 1980s. Dispersion in wage levels actually declined somewhat between 1990 and 2000. This result could be related to the mechanism described above in which the increasing concentration of workers with college degrees may depress the returns they receive.

If true, however, why would college-educated workers continue to flock to labor markets with large populations and stocks of highly educated workers? Glaeser (1999) suggests that workers with a college degree may seek to surround themselves with other college graduates because they are able to learn from one another. Highly educated workers, according to this line of reasoning, are especially committed to the acquisition

[27] A similar argument relating to the relative strengths of the college attainment rate, population, and average wage levels could be made in explaining the divergence patterns of human capital and log population. In those cases, evidently, the mechanisms leading to divergence were stronger than any effects that wage levels might have had.

of productive skills. Since previous work suggests that there may be productive externalities associated with the presence of college-educated individuals (e.g. Moretti (2004)), the positive association between initial college attainment rates and subsequent changes in these rates may reflect the desire of highly educated workers to reside in environments that facilitate learning.

Peri (2002) echoes this view, suggesting that, if skill acquisition is an important reason for the concentration of human capital in cities, we should expect to see large numbers of 'young' college-educated workers in cities. Young workers, after all, are more likely than their older counterparts to seek learning opportunities since they are in the early stages of their careers and, therefore, know relatively little. The evidence he reports is certainly consistent with this idea. Between 1970 and 1990, the ratio of college-educated workers with fewer than 20 years of work experience to those with more than 20 years rose from 1.5 to 2.12 within the metropolitan areas of the US.

Populous cities may also help facilitate the job search process for highly educated married couples. Costa and Kahn (2000) suggest that 'power couples' (i.e. those in which both partners have a bachelor's degree or more) have increasingly moved into large metropolitan areas over the past several decades because cities are more likely to offer job opportunities for both spouses. Large cities, therefore, may provide a solution to the occupational co-location problem.

An additional possibility that deserves to be mentioned involves the amenity value of college-educated workers themselves. That is, while the college-educated may be enticed to locate in cities with a large presence of certain types of establishments (say, eating and drinking places and elementary schools), they might also want to be around other college-educated workers because they desire homogeneity in their social interactions. So, even though concentrations of highly educated workers may be associated with diminishing returns and lower earnings (at least, all else equal), college graduates may still want to surround themselves with other highly educated workers because they find them to be desirable neighbors. Of course, a strong presence of college-educated workers may also be associated with characteristics which have not been accounted for directly here (e.g. low crime, greater civic engagement, good schools), but which are especially desirable to these types of workers.

Whatever the reason, divergence in the distribution of human capital may, over longer time horizons, begin to lead to divergence in earnings and productivity across cities and regions. Although there is only limited evidence of that occurring between 1980 and 2000, it may occur to a greater extent in future decades. In particular, if technologies respond to the supply of skills (e.g., as suggested by Acemoglu (1998)), cities with large stocks of college-educated labor may increasingly utilize technologies that complement these types of workers. This trend may further reinforce the divergence of human capital by encouraging highly educated workers to congregate in the most educated cities as well as lead to greater productivity differentials between cities with small stocks of human capital and those with large stocks.

Table 1: Education Premia by Metropolitan Status

Variable	1980	2000
No High School	1.84 (0.004)	1.73 (0.006)
Some High School	1.96 (0.003)	1.81 (0.005)
High School	2.09 (0.003)	1.98 (0.004)
Some College	2.15 (0.003)	2.11 (0.004)
College or More	2.35 (0.003)	2.38 (0.004)
No High School-Metro	0.013 (0.004)	-0.014 (0.007)
Some High School-Metro	0.031 (0.004)	0.035 (0.005)
High School-Metro	0.046 (0.003)	0.053 (0.004)
Some College-Metro	0.081 (0.004)	0.107 (0.004)
College-Metro	0.156 (0.004)	0.215 (0.004)

Note: Coefficients from regressions of log hourly wages on education indicators and their interactions with a metropolitan status dummy. 1850727 observations for the year 1980; 2135811 observations for the year 2000. Standard errors appear in parentheses.

Table 2: Metropolitan Area Summary Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
Estimated Return	2.47	0.115	2.05	2.84
Population	888590.6	1912186	100376	19397717
Density	578.3	1178.4	6.01	16258.1
College Fraction	0.218	0.065	0.09	0.455
Fraction Agriculture, Forestry, Fisheries	0.007	0.006	0.001	0.063
Fraction Mining	0.006	0.016	0	0.148
Fraction Construction	0.068	0.017	0.033	0.19
Fraction Nondurable Manufacturing	0.077	0.047	0.014	0.365
Fraction Durable Manufacturing	0.123	0.072	0.009	0.453
Fraction Transportation	0.045	0.014	0.018	0.152
Fraction Communications	0.015	0.006	0.004	0.052
Fraction Utilities	0.015	0.007	0.003	0.075
Fraction Wholesale Trade	0.044	0.013	0.015	0.126
Fraction Retail Trade	0.163	0.022	0.096	0.24
Fraction Finance, Insurance, Real Estate	0.061	0.021	0.027	0.24
Fraction Business and Repair Services	0.051	0.018	0.016	0.149
Fraction Private Household Services	0.005	0.003	0	0.02
Fraction Personal Services	0.024	0.015	0.012	0.23
Fraction Entertainment and Recreation Services	0.012	0.012	0.003	0.139
Fraction Medical Services	0.091	0.024	0.04	0.292
Fraction Educational Services	0.105	0.031	0.052	0.27
Fraction Social Services	0.013	0.004	0.005	0.034
Fraction Other Professional Services	0.02	0.009	0.004	0.088
Fraction Public Administration	0.055	0.031	0.015	0.255

Note: Summary statistics taken over 661 city-year observations.

Table 3: Human Capital Accumulation Regression Results

Variable (initial value)	*I*	*II*	*III*
Estimated Return	0.018 (0.011)	-0.02* (0.006)	-0.03* (0.014)
Log Population	0.006* (0.001)	0.003* (0.001)	0.004* (0.002)
Log Density	0.007* (0.001)	0.003* (0.001)	0.002 (0.002)
College Fraction	0.16* (0.015)	0.12* (0.02)	0.11* (0.04)
Fraction Agriculture, Forestry, Fisheries	-0.38* (0.13)	--	0.003 (0.17)
Fraction Mining	-0.13* (0.03)	--	-0.01 (0.06)
Fraction Construction	-0.002 (0.07)	--	0.18* (0.09)
Fraction Nondurable Manufacturing	-0.06* (0.02)	--	0.02 (0.04)
Fraction Durable Manufacturing	-0.03* (0.017)	--	0.03 (0.04)
Fraction Transportation	0.07 (0.08)	--	0.01 (0.08)
Fraction Communications	0.94* (0.19)	--	0.02 (0.23)
Fraction Utilities	-0.29* (0.15)	--	-0.08 (0.15)
Fraction Wholesale Trade	0.01 (0.09)	--	-0.06 (0.1)
Fraction Retail Trade	-0.13* (0.06)	--	-0.01 (0.05)
Fraction Finance, Insurance, Real Estate	0.36* (0.06)	--	0.19* (0.07)
Fraction Business and Repair Services	0.36* (0.11)	--	-0.15 (0.1)
Fraction Private Household Services	-0.56 (0.42)	--	-0.3 (0.4)
Fraction Personal Services	0.02 (0.05)	--	0.06 (0.08)
Fraction Entertainment and Recreation Services	0.04 (0.08)	--	0.004 (0.1)
Fraction Medical Services	0.03 (0.05)	--	0.03 (0.06)
Fraction Educational Services	-0.0005 (0.05)	--	-0.04 (0.06)
Fraction Social Services	0.76* (0.43)	--	0.25 (0.4)
Fraction Other Professional Services	0.83* (0.24)	--	0.1 (0.2)
Fraction Public Administration	0.09* (0.03)	--	0.04 (0.05)

Note: Dependent variable is change in college fraction 1980-90 and 1990-2000. Region indicators and a dummy for the 1980-90 decade appear in all regressions. Column *I* reports coefficients from separate regressions for each regressor. Columns *II* and *III* report coefficients from regressions which include all regressors for which estimates are reported. Heteroskedasticity-consistent standard errors, adjusted for correlation within metro areas appear in parentheses. * denotes significance at 10 percent or better.

Table 4: College Attainment by Major Industry

Industry	1980	1990	2000
Agriculture, Forestry, Fisheries	0.154	0.16	0.154
Mining	0.138	0.179	0.141
Construction	0.072	0.094	0.089
Nondurable Manufacturing	0.112	0.153	0.196
Durable Manufacturing	0.111	0.158	0.183
Transportation	0.09	0.123	0.144
Communications	0.146	0.231	0.328
Utilities	0.123	0.181	0.194
Wholesale Trade	0.152	0.199	0.212
Retail Trade	0.092	0.116	0.136
Finance, Insurance, Real Estate	0.227	0.306	0.364
Business and Repair Services	0.2	0.255	0.33
Private Household Services	0.033	0.052	0.068
Personal Services	0.067	0.105	0.12
Entertainment and Recreation Services	0.194	0.226	0.259
Medical Services	0.219	0.289	0.33
Educational Services	0.546	0.55	0.562
Social Services	0.359	0.41	0.467
Other Professional Services	0.467	0.53	0.537
Public Administration	0.252	0.298	0.352

Note: Fractions of each industry's total employment with a bachelor's degree or higher.

Table 5: Robustness Checks

Variable	*I*	*II*	*III*
Initial Estimated Return	-0.019* (0.01)	-0.028* (0.01)	-0.028* (0.01)
Initial Log Population	0.004* (0.002)	0.004* (0.002)	0.004* (0.002)
Initial Log Density	0.002 (0.001)	0.0006 (0.002)	-0.0001 (0.001)
Initial College Fraction	0.1* (0.02)	0.13* (0.02)	0.11* (0.02)
Initial Eating and Drinking Places Per Capita	12.3* (3.96)	--	6.5 (4.1)
Initial Movie Theaters Per Capita	35.9 (70.1)	--	22.7 (71.1)
Initial Live Entertainment Venues Per Capita	28.5 (34.1)	--	15.3 (28.2)
Initial Museums, Botanical Gardens, Zoos Per Capita	57.1 (150.3)	--	122.6 (141.9)
Initial Elementary Schools Per Capita	77* (40.5)	--	91.9* (39.4)
Initial Colleges and Universities Per Capita	133 (124.2)	--	144.9 (112.9)
Initial Hospitals Per Capita	-216.1* (130.2)	--	-111.3 (102.8)
Δ Fraction Mining	--	0.37* (0.22)	0.32 (0.24)
Δ Fraction Construction	--	-0.01 (0.22)	-0.05 (0.24)
Δ Fraction Nondurable Manufacturing	--	0.03 (0.22)	-0.009 (0.24)
Δ Fraction Durable Manufacturing	--	0.06 (0.22)	0.01 (0.23)
Δ Fraction Transportation	--	0.05 (0.24)	0.06 (0.27)
Δ Fraction Communications	--	-0.29 (0.29)	-0.31 (0.31)
Δ Fraction Utilities	--	0.09 (0.28)	0.04 (0.29)
Δ Fraction Wholesale Trade	--	0.24 (0.25)	0.2 (0.27)
Δ Fraction Retail Trade	--	0.05 (0.22)	0.02 (0.24)
Δ Fraction Finance, Insurance, Real Estate	--	0.56* (0.25)	0.5* (0.27)
Δ Fraction Business and Repair Services	--	0.56* (0.26)	0.55* (0.27)
Δ Fraction Private Household Services	--	-0.04 (0.5)	0.06 (0.53)
Δ Fraction Personal Services	--	-0.11 (0.21)	-0.14 (0.23)
Δ Fraction Entertainment and Recreation Services	--	0.007 (0.24)	-0.07 (0.26)
Δ Fraction Medical Services	--	0.28 (0.24)	0.23 (0.26)
Δ Fraction Educational Services	--	0.48* (0.23)	0.41 (0.25)
Δ Fraction Social Services	--	0.93* (0.38)	0.75* (0.4)
Δ Fraction Other Professional Services	--	0.63* (0.36)	0.58 (0.36)
Δ Fraction Public Administration	--	-0.02 (023)	-0.07 (0.25)

Note: Dependent variable is change in college fraction 1980-90 and 1990-2000. Region indicators and a dummy for the 1980-90 decade appear in all regressions. Heteroskedasticity-consistent standard errors, adjusted for correlation within metro areas appear in parentheses. * denotes significance at 10 percent or better.

Table 6: Residual College Fraction Regressions

Variable (initial value)	*I*	*II*
Estimated Return	-0.026* (0.01)	-0.027* (0.01)
Log Population	0.003* (0.001)	0.003* (0.001)
Log Density	0.001 (0.001)	0.0004 (0.001)
College Fraction	0.08* (0.02)	0.07* (0.02)
Eating and Drinking Places Per Capita	--	5.97* (3.3)
Movie Theaters Per Capita	--	18.7 (64.9)
Live Entertainment Venues Per Capita	--	42 (25.7)
Museums, Botanical Gardens, Zoos Per Capita	--	111.8 (118.1)
Elementary Schools Per Capita	--	64.8* (33)
Colleges and Universities Per Capita	--	159 (99.8)
Hospitals Per Capita	--	-171.2* (98.2)

Note: Dependent variable is the change in the difference between a city's college fraction and its predicted college fraction based on its detailed industrial composition. Region indicators and a dummy for the 1980-90 decade appear in all regressions Heteroskedasticity-consistent standard errors, adjusted for correlation within metro areas appear in parentheses. * denotes significance at 10 percent or better.

Table 7: Growth Regressions

Dependent Variable	Specification	Initial College Fraction	Initial Log Population	Initial Average Log Hourly Wage
Population Growth	*I*	0.28* (0.13)	--	--
	II	--	0.011* (0.005)	--
	III	0.41* (0.14)	0.02* (0.006)	-0.33* (0.08)
Average Hourly Wage Growth	*I*	0.26* (0.04)	--	--
	II	--	--	-0.11* (0.035)
	III	0.42* (0.05)	0.02* (0.003)	-0.39* (0.05)

Note: Regressions of metropolitan area-level population growth and average hourly wage growth on initial values of the college fraction, log population, and average log hourly wages. Region indicators and a dummy for the 1980-90 decade appear in all regressions. Heteroskedasticity-consistent standard errors, adjusted for correlation within metro areas appear in parentheses. * denotes significance at 10 percent or better.

Table 8: Features of the Education, Population, and Average Hourly Wage Distributions

Variable	Statistic	1980	1990	2000
College Fraction	Mean	0.178	0.219	0.253
	10^{th} Percentile	0.122	0.148	0.17
	25^{th} Percentile	0.15	0.18	0.206
	50^{th} Percentile	0.175	0.216	0.248
	75^{th} Percentile	0.203	0.25	0.284
	90^{th} Percentile	0.238	0.296	0.339
	Standard Deviation	0.043	0.054	0.064
	90-10 Difference	0.116	0.148	0.168
	90-50 Difference	0.063	0.08	0.091
	50-10 Difference	0.053	0.068	0.077
	75-25 Difference	0.053	0.069	0.078
Log Population	Mean	12.96	13.06	13.19
	10^{th} Percentile	11.76	11.79	11.91
	25^{th} Percentile	12.07	12.17	12.36
	50^{th} Percentile	12.77	12.88	13.01
	75^{th} Percentile	13.61	13.71	13.9
	90^{th} Percentile	14.37	14.62	14.76
	Standard Deviation	1.057	1.075	1.092
	90-10 Difference	2.61	2.83	2.86
	90-50 Difference	1.6	1.74	1.75
	50-10 Difference	1	1.09	1.1
	75-25 Difference	1.54	1.54	1.54
Average Log Hourly Wage	Mean	2.46	2.45	2.56
	10^{th} Percentile	2.33	2.32	2.43
	25^{th} Percentile	2.39	2.38	2.5
	50^{th} Percentile	2.46	2.45	2.55
	75^{th} Percentile	2.51	2.51	2.61
	90^{th} Percentile	2.58	2.58	2.68
	Standard Deviation	0.098	0.109	0.105
	90-10 Difference	0.24	0.26	0.25
	90-50 Difference	0.12	0.14	0.13
	50-10 Difference	0.13	0.13	0.12
	75-25 Difference	0.12	0.13	0.11

Note: Statistics based on 188 metropolitan areas for the college fraction and average log hourly wage; 187 metropolitan areas for log population.

Appendix

Census Data

The data are taken from the 1980, 1990, and 2000 5 percent samples of the Integrated Public Use Microdata Series (IPUMS) described by Ruggles et al. (2004). Specifically, I use the 1980 5 Percent State Sample, the 1990 5 Percent State Sample, and the 2000 5 Percent Sample. These files have roughly 11.3 million, 12.5 million, and 14 million observations respectively.

To calculate educational attainment distributions among each metropolitan area's labor force, I focus on the working age population (i.e. those who are between the ages of 18 and 65) who report positive wage and salary earnings and are not in school at the time the Census was taken. In estimating the returns to various levels of formal schooling, I further limit the analysis to white males who earn between $1 and $250 per hour. This trimming procedure is designed to eliminate the influence of outlier observations which occasionally appear due to the computation of hourly earnings as the ratio of annual wage and salary earnings to the product of weeks worked and usual hours per week. All dollar figures are converted to real terms (year 2000 dollars) using the Personal Consumption Expenditure Chain Type Price Index.

Potential experience is calculated as the maximum of (age minus years of education minus 6) and 0. Because years of education is not reported for all individuals in the 1990 and 2000 Census, where educational attainment is sometimes reported as a range, I have imputed years of schooling completed using Table 5 of Park (1994).

As noted in the text, metropolitan areas are defined as metropolitan statistical areas (MSAs), New England County Metropolitan Areas (NECMAs), or consolidated metropolitan statistical areas (CMSAs) if an MSA or NECMA belongs to a CMSA. Although somewhat large when considering local labor markets (e.g. New York-Northern New Jersey-Long Island), CMSAs greatly facilitate the construction of geographic areas that have reasonably consistent geographic boundaries. Due to changes in geographic definitions, residents of one MSA within a CMSA in a particular year are sometimes categorized as residing in a different MSA (within the same CMSA) in another year. Aggregating to the CMSA level eliminates any problems arising from this type of definitional change.

Across the 210 metro areas identified in the 1980 Census, the average number of individual level observations used to construct the college attainment and industry share statistics is 14484.1 (minimum = 1557, maximum = 315128). Among the 206 metro areas identified in the 1990 data, the average is 15863.4 (minimum = 1426, maximum = 329632). In the 245 metro areas from the 2000 Census, the mean is 16526.4 (minimum = 1426, maximum = 371278). When confining the sample to white males only for the college return regressions, the mean numbers of observations (minimum, maximum) per metro area are 6928.1 (666, 144886) for 1980, 7275.7 (670, 144698) for 1990, 6600.7 (516, 134898) for 2000.

A complete list of the detailed industries appears in the IPUMSs documentation at www.ipums.org. These can be found in the link to the industry codes for 1980. The 1990 and 2000 codes are translated into equivalent 1980 codes using the programs that accompany this paper.

Additional Data Details

Metropolitan area population density is calculated as a weighted average of county-level densities where the weights are given by population shares. This measure is intended to give a better sense of the average density per square mile faced by a typical city dweller than average density (metropolitan area population to metropolitan area land area) which may be misleading, particularly among cities in the western US which encompass extremely large, but sparsely populated counties.

US Census Regions

West: Washington, Oregon, California, Nevada, Idaho, Montana, Wyoming, Utah, Colorado, Arizona, New Mexico, Alaska, Hawaii

Midwest: North Dakota, South Dakota, Nebraska, Kansas, Minnesota, Iowa, Missouri, Wisconsin, Illinois, Michigan, Indiana, Ohio

Northeast: Maine, Vermont, New Hampshire, Massachusetts, Rhode Island, Connecticut, New York, Pennsylvania, New Jersey

South: Texas, Oklahoma, Arkansas, Louisiana, Kentucky, Tennessee, Mississippi, Alabama, West Virginia, Delaware, Maryland, District of Columbia, Virginia, North Carolina, South Carolina, Georgia, Florida

References

Acemoglu, Daron. (1998) "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." Quarterly Journal of Economics. 113 (4), 1055-1089.

Barro, Robert. (1991) "Economic Growth in a Cross Section of Countries." Quarterly Journal of Economics. 106 (2), 407-443.

Barro, Robert and X. Sala-i-Martin. (1992) "Convergence." Journal of Political Economy. 100 (2), 223-251.

Barro, Robert and X. Sala-i-Martin. (1995) Economic Growth. New York: McGraw-Hill.

Black, Duncan and Vernon Henderson. (1999) "A Theory of Urban Growth." Journal of Political Economy. 107 (2), 252-284.

Card, David. (1999) "The Causal Effect of Education on Earnings." In Handbook of Labor Economics, Vol. 3, O. Ashenfelter and D. Card eds. Amsterdam: Elsevier.

Costa, Dora and Matthew Kahn. (2000) "Power Couples: Changes in the Locational Choice of the College Educated, 1940-1990." Quarterly Journal of Economics. 115 (4), 1287-1315.

Dee, Thomas. (2004) "Are There Civic Returns to Education?" Journal of Public Economics. 88 (9-10), 1697-1720.

Durlauf, Steven and Danny Quah. (1999) "The New Empirics of Economic Growth." In Handbook of Macroeconomics, Vol. 1A, J. Taylor and M. Woodford eds. Amsterdam: Elsevier.

Glaeser, Edward. (1999) "Learning in Cities." Journal of Urban Economics. 46, 254-277.

Glaeser, Edward, Jed Kolko, and Albert Saiz. (2001) "Consumer City." Journal of Economic Geography. 1 (1), 27-50.

Glaeser, Edward and Albert Saiz. (2003) "The Rise of the Skilled City." Harvard Institute of Economic Research Discussion Paper #2025.

Glaeser, Edward and Raven Saks. (2004) "Corruption in America." Harvard Institute of Economic Research Discussion Paper #2043.

Glaeser, Edward, Jose Scheinkman, and Andrei Shleifer. (1995) "Economic Growth in a Cross-Section of Cities." Journal of Monetary Economics. 36, 117-143.

Lochner, Lance and Enrico Moretti. (2004) "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self Reports." American Economic Review. 94 (1), 155-189.

Milligan, Kevin, Enrico Moretti, and Philip Oreopoulos. (2004) "Does Education Improve Citizenship? Evidence from the United States and the United Kingdom." Journal of Public Economics. 88 (9-10), 1667-1695.

Moretti, Enrico. (2004) "Human Capital Externalities in Cities." In Handbook of Regional and Urban Economics. Vol. 4, J.Vernon. Henderson and Jacques-Francois Thisse eds. New York: Elsevier, 2243-2291.

Park, J. (1994) "Estimation of Sheepskin Effects and Returns to Schooling Using the Old and New CPS Measures of Educational Attainment." Princeton Industrial Relations Section Working Paper # 338.

Peri, Giovanni. (2002) "Young Workers, Learning, and Agglomerations." Journal of Urban Economics. 52 (3), 582-607.

Ruggles, Steven, Matthew Sobek, Trent Alexander, Catherine Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander. (2004) Integrated Public Use Microdata Series: Version 3.0. [Machine-readable database] Minneapolis, MN: Minnesota Population Center [producer and distributor], 2004.

Simon, Curtis and C. Nardinelli. (2002) "Human Capital and the Rise of American Cities, 1900-1990." Regional Science and Urban Economics. 32, 59-96.

US Bureau of the Census. (1999) USA Counties 1998 on CD-ROM. [Machine readable data file]. Washington, D.C.: The Bureau.